

# Hadoop企业级Hadoop平台二次开发-第四天

## 一、课前准备

1. 理解HDFS的启动流程
2. 理解HDFS元数据管理流程

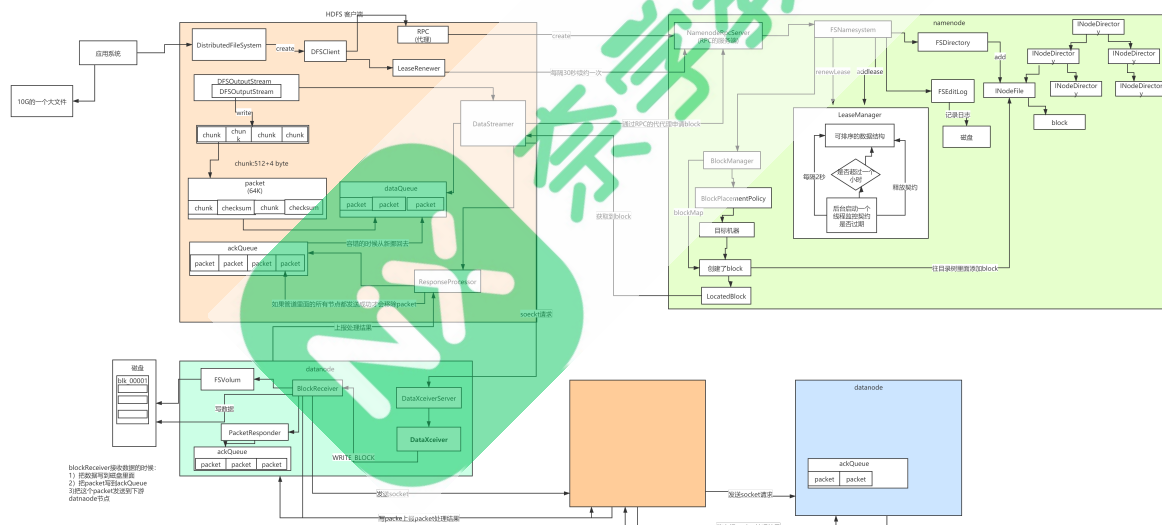
## 二、课堂主题

本次课根据之前课程的学习，掌握HDFS写数据流程的核心架构设计。

## 三、课程目标

1. 学习HDFS写数据的完整流程（HDFS的源码写得不好）
2. 学习HDFS写数据流程的架构设计
3. 为什么要建立数据管理

## 四、知识要点（150分钟）



## 六、扩展

先引入一个小的背景，假如多个客户端同时要并发的写Hadoop HDFS上的一个文件，这个事儿能成吗？明显不可以接受啊，因为HDFS上的文件是不允许并发写的，比如并发的追加一些数据什么。所以HDFS里有一个机制，叫做文件契约机制。

也就是说，同一时间只能有一个客户端获取NameNode上面一个文件的契约，然后才可以向获取契约的文件写入数据。

此时如果其他客户端尝试获取文件契约的时候，就获取不到，只能干等着。

通过这个机制，可以保证同一时间只有一个客户端在写一个文件。

在获取到了文件契约之后，在写文件的过程期间，那个客户端需要开启一个线程，不停的发送请求给NameNode进行文件续约，告诉NameNode：

NameNode大哥，我还在写文件啊，你给我一直保留那个契约好吗？

而NameNode内部有一个专门的后台线程，负责监控各个契约的续约时间。

如果某个契约很长时间没续约了，此时就自动过期掉这个契约，让别的客户端来写。

## 七、互动

1. 为什么HDFS选择要建立数据管道？

